GRANT NUMBER: N00014-94-1-0674

Progress Report: Towards a theory of learning during tutoring

Kurt VanLehn
Michelene T. H. Chi
William Baggett
R. Charles Murray
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260
VanLehn@cs.pitt.edu
(412)624-7458

November 16, 1995

Contents

This progress report has 4 sections. The first gives an overview of the project, our pilot study and our plans for future work. The second and third sections present the results from the pilot study in more detail. The last section discusses our plans for the main study.

1 Executive Overview

1.1 Goals

The goals of the research are basically to (1) understand why tutoring is so much more effective than other forms of instruction, and (2) to understand what exactly causes learning while a student is being tutored. More details on the research objectives can be found in the proposal.

1.2 Methods

Our main method is to take protocols of students as they are being tutored. However, we felt that too much communication for us to record would occur at the non-verbal level if the tutor and student were seated side by side. Moreover, the tutor would hear the verbal protocol of the student, and that would necessarily affect the tutor's tutoring. Lastly, the presence of the tutor would probably affect the students' protocol. Therefore, we decided to put the tutor and student in separate rooms.

The student and tutor communicate in two ways. First, students do all their work on a computer screen, namely, the user interface module to the OLAE system. It allows drawing vectors, drawing coordinate axes and typing equations. The tutor has a duplicate of the screen in his room so he can see the student's work.

Second, the student and tutor communicate by phone. To avoid dialing, the phone line is kept open, but the tutor keeps his phone receiver on the desk when the two participants are not talking so that he cannot hear the student's verbal protocol. When the student types "help" or the tutor sees something he wants to talk to the student about, the tutor picks up the phone and speaks. When they are done talking, the tutor says good-bye and puts the phone receiver back down on the desk. In the student's room, the phone is a speaker phone so that the conversation can be recorded along with the verbal protocol.

The verbal protocol is recorded by PROTO, a digital recorder built by the project. It

captures not only the sound but also synchronizes it with the user interface events. This allows the analyst to play back both sound and screen at the same time. We have found that this makes transcription much faster and more reliable. Moreover, even after the protocols are transcribed, we often go back to the replay the audio-visual source in order to check our interpretations. Exact timing and vocal intonations are often quite powerful for disambiguating competing interpretations.

1.3 Analyses

Our original intention was to do the same types of analyses that were done with the Chi et al. (1989) self-explanation data. We have undertaken 4 of them:

1. Chi et al. (1989) classified utterances made by students and found utterances classified as "self-explanations" correlated with learning. In our pilot study, we classified tutoring episodes, and began to look for correlations with learning effectiveness.

2. Chi and VanLehn (1991) classified self-explanations according to their physics content in order to discover what kinds of reasoning lay behind them. In the current study, we classified tutoring episodes by their physics content.

3. VanLehn and Jones did a series of analyses that developed a model, CASCADE, which could explain most of the protocols (VanLehn et al., 1992; VanLehn and Jones, 1993b). In the current study, we developed a taxonomy of all episodes in the protocols (not just the tutorial ones) that is also intended to lead to a computational model that will explain the protocols.

4. (VanLehn, 1995c) located all the learning events in the Chi protocols in order to determine what caused learning and why self-explanation made learning more effective. In the current study, we collected the learning events in the protocols in order to determine the causes of learning and why tutoring is so effective.

The research goals are the same for both the single-student studies and the present student-tutor studies, namely, to understand how learning works and why some situations are more effective than others. Thus, it makes sense to apply the same analyses in both studies.

1.4 Results so far

Only 2 of the 4 analyses have born fruit so far. The learning-events analysis (conducted by Micki Chi) showed that learning was not caused by a tutor who inferred the deep misconceptions of the student and adapted the instruction or remediation accordingly. Rather, the tutor acted like an Andersonian model-tracer. The student learned by (1) unaided construction of task-specific knowledge, (2) getting immediate feedback from the tutor that conveyed a necessary piece of knowledge, or (3) asking the tutor a direct question whose answer conveyed a piece of missing knowledge.

The effort to classify or parse all episodes in the protocols led to some ideas about coached problem solving (viewed as a special kind of collaborative problem solving) that seem to be adequate for developing a computational model that would cover a very large proportion of the protocols. This should allow us to deepen our understanding of how learning occurs, and particularly, the contexts of its occurrences.

1.5 Plans for further work

There are three main problems with the pilot protocols. We intend to fix them in the next round of experimentation. The problems and their solutions are:

- The user interface (OLAE) is hard to use, and the digital recording equipment kept breaking. We have fixed the digital recorder (the problem was an intermittent fault in the disk drive!), but we do not want to fix OLAE. Consequently, we will include more user-interface instruction before the tutoring.

- It is difficult to detect learning if one does not know what the student knows before and after tutoring. We will give students pre- and post-tests, although developing good ones will be non-trivial.

- When the student is talking, either to the tutor or to the digital recorder alone, we can generally infer what the student is thinking. However, when the tutor is talking, the student can't talk, so we have no way to know what the student is thinking. We tried using a cued retrospective protocol with one student, but were not satisfied with

the results. We may try other students before giving up. We may also ask tutors to avoid lecturing the students and to make their comments as brief as possible.

Although characterizing tutoring strategies is not a goal of this project, we cannot help but notice that there are large differences between tutors. A tutor from the physics department simply lectured—we did not even transcribe her sessions. The PI (KVL) acted like an Andersonian model-tracing tutor. All of the transcribed protocols come from his sessions. The third tutor (Patty Albacete) virtually never interrupted the student, and only gave feedback at the very end of the problem or when asked directly by the student.

The students' opportunities for learning and probably their general learning strategies are affected by the tutor's strategies. In order to avoid an unintentionally parochial view of learning, we need to use a variety of tutors. This creates some analytic difficulties that we still need to think through.

Our major concern at the beginning of analysis was that the students were not talking enough just after they finished a tutorial interaction. Our belief at the time was that this period was the critical one for learning because students would use it to assimilate the tutor's message into their own understanding of physics problem solving. We are now convinced that this particular time (just after the tutorial interaction finishes) is probably less critical than we initially thought. More learning probably occurs *during* the tutorial interaction than immediately afterwards because students often will not end the tutorial interaction until they feel that they know what to do next. In fact, of the 3 learning events uncovered in Chi's analysis, 2 occurred during tutorial interactions and 1 occurred just afterwards.

1.6 Conclusions

Although we initially had doubts about that this experimental method could achieve the goals of the study, analysis of the pilot data has convinced us that the method will actually work. Moreover, we found some interesting initial results that suggest that the theoretical framework (the CASCADE model) used for understanding the self-explanation data (non-interactive instruction) can be extended quite naturally to help us understand learning in the context of interactive instruction (tutoring). These results have provided us with specific hypotheses to be tested in the main study.

2 Parsing the protocols: Towards a model of tutoring

One of the 4 analyses conducted sought to parse the tutoring protocols, that is, to assign a hierarchical interpretation to the whole protocol that relates each section to the others. Since it is far to early to render this in a computational model, the analysis takes the form of a taxonomy of protocol episodes (for an advance organizer, see Figure 1 on page 15). We present our taxonomy top down, starting with the broadest classification.

2.1 Basic types of tutoring activity

The top level of the hierarchy basically separates the bulk of the protocol from various types of uninteresting behavior. The basic categories are:

- *Coached problem solving.* This is the main type. The student and tutor work together in order to solve a problem. Their conversation is almost always about the problem, with only occasional references to generic, problem-independent principles or concepts of physics.

- *Lecturing.* Occasionally the tutor lectured the student. This was almost always directed at a knowledge flaw that the tutor believed that the student possessed which had shown up as an error of some kind. Thus, lectures could be considered part of coached problem solving. But when the coaching becomes really long, then it is more likely (in these protocols) to have little impact on the student. So we classify "long" monologues about generic, problem-independent domain knowledge as lectures. Clearly, there is a fine line here that eventually needs to be clarified.

- *Recap of problem solving.* The tutor and student sometimes review the solution of a problem. In principle, this could be just as rich an experience as coached problem solving, but it did not seem to be in these protocols.

- *Extraneous topics.* Occasionally the students raised topics that were not directly relevant to the problem solving at that moment. For instance, student A asked the tutor about a contradiction between something her teacher had said and something her textbook had said.

- *Extended miscommunication.* Sometimes the tutor misunderstands the student's goals or beliefs to such an extent that they can hardly communicate. This occurred at the end of Student B's protocol of the first problem, where the tutor misdiagnosed B's problem, tried a Socratic repair, and thus totally confused B. The episode lasted for several minutes, and was excruciatingly frustrating for both participants.

- *Floundering.* Sometimes a student gets hopelessly lost and/or confused. However, instead of asking the tutor for help, the student just keeps trying alone. If the tutor finally interrupts (see Coordination below), then the tutor can sometimes resurrect the student from a floundering state by engaging in coached problem solving. However, Extended miscommunication sometimes ensues as well. There are several examples of floundering in B's protocol.

Although all these types of behavior are interesting for one reason or another (except possibly lecturing), coached problem solving is the most common type of activity and probably the one that causes most of the learning. Since our goal is to understand why tutoring is so effective, we further analyzed episodes of coached problem solving and ignored all other types of episodes.

2.2 Types of coached problem solving

Within the classification of coached problem solving, there were basically three types of activities.

- *Progress.* The tutor and the student make progress on solving the problem. Their beliefs, goal structures and terminology agree.

- *Coordination.* Sometimes the tutor cannot tell what the student is doing or what something that the student has written means. The tutor interrupts the student and asks.[1]

- *Debugging.* Someone, usually the tutor, has detected a difference in their mutual beliefs, goals or terminology, and has initiated an episode of debugging or repairing

[1] Actually, we are not happy with breaking out Coordination at this level of the taxonomy. Communication activity (and miscommunication as well) actually is interwoven in all parts of the protocol. We do not yet know how exactly to conceptualize this.

the error.

This section examines each of these activities. Episodes of the first category, Progress, are the most common, so they are examined further in the next section.

2.2.1 Progress

As Newell and Simon discovered back in the 60's, the task itself determines the possible lines of reasoning that a problem solver can take. They called the set of possible lines of reasoning for a particular problem the *problem space* for that problem. Physics problem solving is so well-defined that it is possible to describe the whole problem space for most problems. When the student is following one of the lines of reasoning in the problem space of the current problem, then we classify the episode as Progress.

Because there are two agents capable of reasoning (the student and the tutor), the line of reasoning can be carried forward in multiple ways depending on who actually does the reasoning. In particular, each inference along the line of reasoning can be done one of 5 ways:

1. The student makes the inference without assistance from the tutor. This case includes occasions when the student reaches an impasse but manages to repair it without help from the tutor.

2. The student tries to make the inference, reaches an impasse, and seeks help from the tutor, thus initiating a debugging episode. If the tutor happens to be listening as the student reaches the impasse, the tutor may intervene without being asked for help.

3. The student applies an incorrect piece of knowledge, and the tutor initiates a debugging episode. Sometimes students feel so uncertain about such inferences that they initiate the debugging by asking the tutor to check the line or by telling the tutor that their result doesn't make sense.

4. The tutor makes the inference while the student watches silently.

5. The tutor makes the inference and the student asks a question about it. If the student's question has an incorrect presupposition, the tutor usually initiates a debugging

episode. If not, the tutor answers the question directly. The student may have other follow-up questions, which are handled the same way.

In principle, there could be more than these 5 ways to advance the line of reasoning, but these are major ones we have seen so far.

2.2.2 Coordination

Coordination is a relatively brief activity wherein the tutor finds out what the student is thinking. There are two basic kinds. If a student makes no visible action for several minutes, the tutor interrupts with something like, "Need help?" Sometimes the offer is rejected. If not, then the tutor asks the student what she is working on. This usually leads either to an episode of Progress if the student is on the right tract, or Debugging if not.

The other type of coordination occurs when the student has entered something that the tutor literally does not understand, so he can't determine whether it is correct or incorrect. For instance, the tutor might not know what a certain variable stands for, or how an equation was derived. The tutor interrupts to ask the student what the entry means. If this meaning makes the entry correct, the tutor thanks the student and falls silent. Otherwise, a Debugging episode follows.

2.3 Debugging

Debugging occurs when the student has reached an impasse or made an incorrect inference, and the tutor has been called in to help. In general, debugging episodes have three goals: track down the piece of knowledge that should have been applied, get the student to believe it (if the student doesn't already), and restart the problem solving along a correct line of reasoning. Exactly how these three goals are achieved depends somewhat on the particular error or impasse that initiated the debugging episode.

There are basically three kinds of errors. The most common one is that the student has an incorrect belief or a missing belief. For instance, the student has interpreted the F in $F = ma$ as a single force rather than a sum of forces. The second type of error occurs when the student use an incorrect term to refer to something. For instance, the student might use "weight" to refer to a quantity that is actually a mass. Often the terms are nouns, but

occasionally they are verbs. For instance, student A asked if she should "convert Newtons to kilograms?" The student was referring to an application of $w = mg$ (weight is mass times acceleration due to gravity), which is not a units conversion. The third type of error is when the student adopts a goal that the tutor would not adopt at that point. For instance, students often tried to write equations before completing a force diagram.

Once initiated, a debugging episode has a predictable structure. If necessary, the tutor "backs up" the mutual goal structure to a place where the underlying source of the error seems to be. For instance, if the student does not draw any forces but starts immediately to write an equation that implies that not all the forces in the problem have been identified, then the tutor asks the student to reconsider or even draw the forces on that object. Thus, the student is asked to abandon the goal of writing Newton's law for the object and reopen the goal of identifying forces on the object: a goal that the student presumably finished earlier even though it resulted in no visible actions.

The second phase in debugging, which again might be optional, is when the tutor asks the student questions intended to narrow in on the source of the student's error. For instance, when the tutor saw that student Q assigned an incorrect, positive sign to an object's acceleration, he asked the student which direction the object was accelerating. It turned out that the student had the direction wrong, which in turn caused the sign error.

Once the tutor had localized the error, he would sometimes simply tell the student the correct thing to do. However, he often tried to get the student to construct it instead. For instance, if the student had the direction of the acceleration wrong, the tutor might say, "Oh, so the object is speeding up, is it?" The tutor has used the phrase "speeding up" to co-refer with "positive acceleration," which is a subtle way to get the student to notice a contradiction. Perhaps the most elaborate example of this kind of repair was multi-minute episode wherein student B was induced to identify a friction force that he had overlooked.

As the last act of a debugging episode, the participants may negotiate about what goal to resume the problem solving with. For instance, if the bug is that a force was not identified, the tutor may insist that the student draw a force diagram, or the tutor may allow the student to continue writing the equation that prompted the debugging episode.

2.4 Types of Progress

During episodes of Progress, either the student, the tutor or both make inferences that move forward along a line of reasoning. We earlier classified episodes of Progress by who made the inferences. One could call this the "agency" taxonomy. In this section, we classify them by their content. As it turns out, the content taxonomy interacts with the agency taxonomy.

There are basically 3 types of physics inferences used in solving a problem: qualitative analysis, equation generation and planning of equation generation. Each is discussed separately below.

2.4.1 Qualitative Analysis

During qualitative analysis, there is no ordering on the production of inferences. Moreover, unless the student is drawing forces as she goes, the tutor has no entries to see. This makes it nearly impossible for the tutor to follow the student as she does qualitative analysis. In the protocols, the tutor typically left the student alone while she was doing qualitative analysis, but if the qualitative analysis turned out later to have errors, the tutor would go back and redo the qualitative analysis with the student.

Doing the qualitative analysis with the student typically began with the tutor asking the student about all inferences that the tutor knew were important and had not yet been demonstrated in the student's work. For instance, one student had drawn a force and a velocity for a body, but had not draw the other force acting on the body nor the body's acceleration. When she later misapplied Newton's law by leaving out a force, the tutor reopened the qualitative analysis phase. He first asked the student what the drawn force was. Although strictly speaking, this step was not necessary, it ensured that the student and the tutor agreed on the meaning of the vectors. After the student responded correctly, the tutor asked if there were any other forces. The student didn't see any, so the tutor reminded her of one that was mentioned in the problem statement. After the student drew this force, the tutor moved on to the next important set of inferences, namely those involving the existence and direction of acceleration. He asked the student about each of these, although

he did not insist that she draw the acceleration on the diagram. In the process, he made sure that the student saw the relationship between the object's acceleration, its initial velocity and its final velocity. Thus, the tutor covered all the important inferences in the qualitative analysis, mostly by asking the student about them.

Terminology is especially important during the qualitative analysis phase. The tutor would often deliberately refer to forces by their proper names (e.g., the force of gravity on an object is called its "weight") even when the student seemed not to know the technical meaning of the terms. This is a subtle way of teaching the meanings of the phrases. Sometimes this tactic would provoke the student to ask about the terminology. Often it did not, and served only to alert students that something about the term was strange; they would begin to use it in an ambiguous manner. Lastly, students would sometimes use canonical terms incorrectly, in which case the tutor would correct them. For instance, one student incorrectly referred to a force as a "normal force," probably because it was directed upward. The tutor explained that normal forces are due to surfaces.

One way to model qualitative analysis is as the production of a set of important propositions. Sometimes the student does the inference, and may tell the tutor about it. If the student does it incorrectly, the tutor will often debug the inference. Sometimes the tutor does the inference and tells the student about it. In this case, the student may ask for clarification. Sometimes the student may even ask about the proposition before the tutor produces it (e.g., after the tutor points out that an object is acceleration, the student may ask which direction the acceleration is headed).

2.4.2 Equation generation

By equation generation, we mean the basic means-ends analysis (MEA) production and solution of a set of equations. The idea is to pick a goal if you haven't got one already, pick an operator that will help achieve the goal, subgoal recursively to satisfy the prerequisites of the operator as necessary, then apply the operator. This cycle repeats until there are no more goals left to achieve. In the case of equation generation, most goals concern finding values for variables. The operators are typically equation-based ones, such Newton's law.

As discussed in the agency taxonomy, each step of the MEA procedure can be done by

either the student or the tutor. If the student does it, it can be done either correctly (in which case the tutor usually says nothing) or incorrectly, in which case the tutor usually debugs the errors. Sometimes the student tries to do the step, but reaches an impasse and may ask for help (or just flounder silently until the tutor interrupts—see Coordination above). Thus, there are 3 student-control cases (a) do it alone, (b) do it incorrectly and get debugged by the tutor, or (c) fail to do it and get help with the impasse from the tutor.

If the tutor does the step instead of the student, the student may or may not ask a question about the step. This distinction is reflected in agency taxonomy (see Figure 1). The tutor probably does the step because he believes that the student could not do it successfully. For instance, it seems clear in several cases that the tutor insisted on drawing a force diagram when the students could not figure out why. None objected to it, and none even questioned the tutor. After that, most students started to draw force diagrams themselves. When the tutor takes a step, this is a form of scaffolding. As discussed later, it can cause learning.

If the tutor had introduced an operator and set the student to working on its first subgoal, then the tutor would often take over again when the student had finished the subgoal, and set the student to working on the next subgoal of the tutor's operator. In such cases, the tutor either had not explained the operator when introducing it or felt that the explanation had not been understood. Thus, he broke the operator into parts (subgoals) and coached the student through the process of applying it.

2.4.3 Planning the equation generation

Before discussing the episodes where planning occurs, it is worth characterizing planning in this task domain.

Suppose the problem is "A 6 N force pushes rightward and horizontally on two adjacent blocks sliding on a frictionless plane. The left block has mass 2 kg and the right block has mass 6 kg. What is the force exerted by the small block on the large block?" A plan for equation generation might sound something like this:

> First we'll apply Newton's law to both blocks together and get their acceleration. We can do that because they give us the force acting on them both

and their mass. Then we'll apply Newton's law to the large block, and use it to calculate the force acting on it.

The defining characteristic of a plan for equation generation is that it does not mention the algebraic details nor even the geometric details. For instance, this plan does not mention the axis along which Newton's law will be applied. Experts are often able to generate such plans with only a short study of the problem, while only the best students can do so (Chi et al., 1982).

Plans are optional. One can simply generate equations without first generating a plan for their generation. Experts use planning for several reasons. For instance, the branching factor for searching in the space of abstract equations (the ones used in plans) is less than the branching factor for searching in the space of equations. Novices probably cannot generate plans because they do not know the abstract equations well enough. Thus, they cannot hold the partial plans in memory during the search (VanLehn, 1995a).

In our protocols, planning was not common. Some students planned by themselves, especially on the two-block problem used as an illustration above. The tutor would sometimes try to tell a plan to the student, especially for the last, most complex problem. The students never seemed to get it. Apparently, communication failed because of complexity in either the planning operators themselves, the vocabulary for describing them or the mental computation required to construct and verify a plan from its verbal description.

2.5 Towards a model of Coached Problem Solving

Figure 1 shows the taxonomy of episode types that we have devised in order to cover these protocols. It is a little complex in that the category of Progress is cross-categorized, once by who does the inference (the agency classification) and once by the content of the inferences.

As mentioned earlier, the most popular category is Coached Problem Solving. We estimate that it accounts for about 80% to 90% of the protocol episodes, except in student B's protocol where there is a great deal of Floundering and Extended Miscommunication.

The taxonomy under the Coached Problem Solving node is the beginning of a computational model. This model is the main objective of the AASERT project that is associated with this one. In order to move the analysis closer towards specifying the model in more

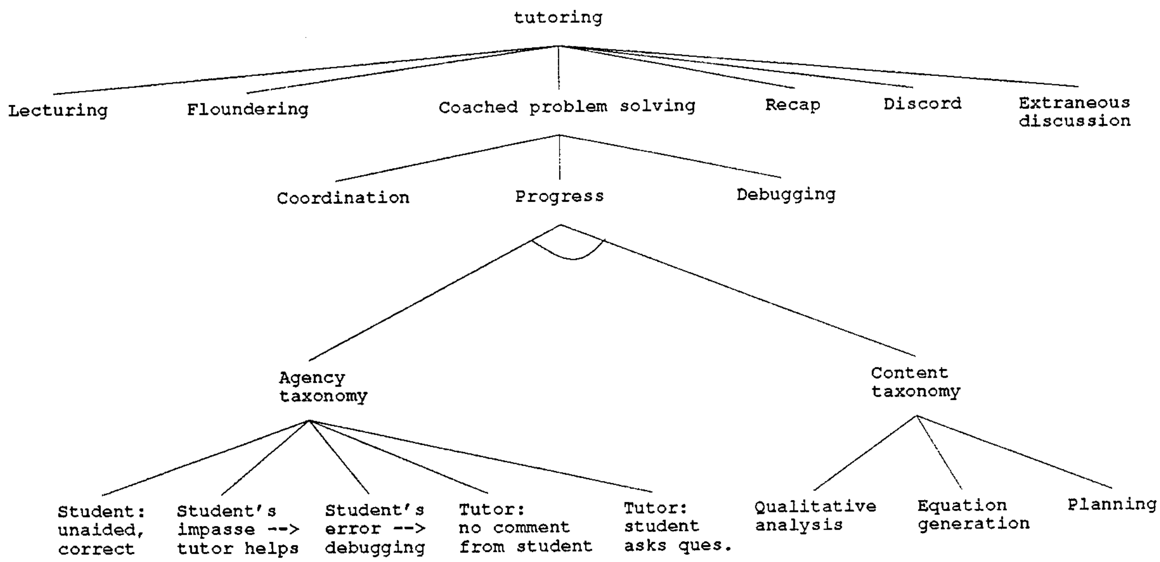

Figure 1: A taxonomy of tutoring episodes

detail, we need to analyze the protocols in typical Newell and Simon (1972) style. Chas Murray is working on this. However, as a down payment, an appendix shows a rough draft of an analysis of the protocol of student Q doing one problem. It illustrates many of the categories of activity described above, as well as illustrating what we mean by "parsing" a protocol.

The amazing thing about the model is its simplicity. In its basic structure, it is just a problem solver of the CASCADE style, with a STEPS-like debugger added (Ur and VanLehn, 1995) and something for Coordination.

The surprise (and the simplicity) is that it appears to only take two modifications to create a model of coached problem solving. First, we need to allow either the tutor or the student to take any of the inferences, and to do so in a variety of ways (i.e., the agency classification). The second change is that the mere act of communication can break down in that the tutor can catch errors in the student's use of terminology, or the student can question the tutor's use of terminology.

In principle, the model could have been much more complicated. For instance, peer problem solvers often spend a great deal of time negotiating what to do next and whether an inference is correct (e.g., Baker, 1994). Negotiation appears to play very little role in these protocols, so it has been omitted from the model for parsimony. The thrust of the

coverage claim is not that other dyadic educational activities (e.g., negotiation; telling long stories; elaborate reflective follow-up; etc.) never occur, but that they are not very common in tutorial dialogs. Thus, to the extend that learning does occur during tutoring, it must be due to the relatively simple activities encompassed by the model. Thus, the simplicity of the model encourages us to think that we can ultimately deeply understand why tutoring is such an effective pedagogy.

3 Where's the learning?

We can approach this question (where is the learning, what causes it, etc.) both from the bottom up (by analyzing existing cases of learning) or from the top down (by analyzing the opportunities for learning implicit in the model). The first section is a bottom up analysis; the second is a top down one.

3.1 Learning events

As the ultimate goal is to find out why tutoring is so effective, a good first step is to examine where learning occurs and determine what caused it. This was the main thrust of an analysis by Micki Chi, who joined the CASCADE project for a month. Her report is attached. Since it was written as an internal memo for the CASCADE group, and contains many details (e.g. line numbers) that aren't germane here, a summary and discussion in provided.

Briefly put, Chi concluded that during the one problem that she analyzed most intensely, one of the students learned three pieces of knowledge. Although Chi does not use the term "learning events," that is what they are.

One learning event was the discovery that mass and weight are distinct. It occurred when the student was working alone. Although the tutor had prepared the ground by hinting that mass might be problematic, the tutor did not actually tell the student this piece of knowledge. The student seems to have constructed it herself. Similar autonomous learning events have been observed in physics problems solving (VanLehn, 1995c), puzzle solving (VanLehn, 1991) and counting (Siegler and Jenkins, 1989). As noted in (VanLehn and Jones, 1993a), for such learning events to occur, it is important that the learner be

on a correct solution path. If they are not on such a path, it is likely that they will learn incorrect beliefs. Moreover, if the learner is not confident that she is on a correct path, then she may have no confidence that her discoveries are correct. The fact that the tutor keeps a student on the correct path, and that the student believes that the tutor does so, may contribute to the effectiveness of learning via this type of learning event.

During the second learning event, the student learned that weight is a kind of force by asking the tutor a direct question, "but why don't I need to take the, um, the gravity into account, thought?" The tutor explains that the object's weight, which the student has discussed with the tutor, *is* the force due to gravity on the object. Clearly, this kind of just-in-time inferencing is unique to tutoring. All the constituents for forming the new piece of knowledge were active in the student's working memory. The tutor just needed to form the connection. Clearly, this kind of learning event is part of the secret of tutoring's effectiveness.

During the third learning event, the student learns that when an object is speeding up, its acceleration is in the direction of the movement, and conversely, when an object is slowing down, its acceleration is opposed to the direction of movement. The student had written an equation with the wrong sign for acceleration. The tutor asked her if the sign was correct. The student waffled. The tutor asserted that if the body is accelerating downward and moving downward, "that would mean [it is] speeding up." This is a case of the student accepting an assertion, and generalizing it. However, the student was perhaps more receptive to this assertion than otherwise because she had just tried and failed to make the correct inference. Again, this kind of learning event is perhaps unique to tutoring.

This kind of learning events analysis could be carried forwards by encoding the other parts of the protocols. However, as discussed later, it would be better to obtain new protocols where evidence for the learning events is more clear.

3.2 Learning predicted by the model

Essentially, the model is just a problem solver, albeit one that lets two agents take turns in implementing its interpretation steps. A problem solver is just a means of *using* knowledge to solve problems. The key idea is that learning comes via the *use* of knowledge, but in

several forms.

In the problem solver, there are basically 5 ways to make an inference (the agency taxonomy), and each can cause a different kind of learning event:

1. The student makes the inference, employing the knowledge if it exists and constructing it from overly general rules if it doesn't. This is exemplified by the first of Chi's learning events, and many others from the literature.

2. The student can try to make the inference, reach an impasse, and get help from the tutor. If the tutor's help makes sense to the student (and the student may have to ask some follow-up questions to insure this), then the student can learn a new piece of knowledge, thus filling a gap in the student's knowledge.

3. The student can apply an incorrect piece of knowledge, and get debugged by the tutor. The student may have to lend a hand to the debugging effort in order for the remediation to be successful. The third Chi learning event is an illustration of this type of learning event.

4. The tutor can make an inference while the student watches silently. If the knowledge used by the tutor can be easily inferred by watching, and the student doesn't have that knowledge, then the student can learn the piece of knowledge. However, we suspect that this rarely occurs because the tutor moves along so quickly that the student may not have the time to do the requisite inductive inferences. An exception might be when the tutor has already explained the knowledge once, and is walking the student through its application.

5. The tutor makes an inference and the student asks a question about it. This is illustrated by the second Chi learning event. The tutor was about to go on, thus implying that all the forces had been drawn on the force diagram. The student questioned this step. Interestingly, the tutor did not directly answer the student's question. The question was based on a presupposition, which is that the gravitational force was not already drawn on the diagram. The tutor inferred the buggy knowledge from the presupposition and fixed it with a simple statement. In general, learning is

probably more often due to simple, direct answers to questions.

3.3 How could tutoring be made more effective?

Once one has a model of how learning *could* occur, it is relatively simple to determine policies that will increase the amount of learning that will occur. Some would be policies or strategies that students would carry out; others would be the responsibility of the tutor. This section discusses a few such policies. Each such policy is a hypothesis about what effective student-tutor dyads do differently from ineffective dyads. Such a policy is also a suggestion for meta-instruction: teaching a student how to learn more effectively and teaching a tutor (or tutoring system) how to tutor more effectively. Thus, these policies are like the meta-strategy of self-explaining examples.

One strategy that students can apply is to be as constructive as possible while the tutor is scaffolding the problem solving. For instance, if the tutor demonstrates an inference, then students should decide whether they would make the same inference. If not, then they should try to construct the knowledge to make that inference. This might lead them to ask the tutor a question about the tutor's inference.

Bugs can hide beneath the scaffolding. That is, the tutor should avoid scaffolding the student as much as possible because it may be that if the student were allowed to produce the inference instead, the tutor might observed some incorrect knowledge (bug) that could then be remedied.

Both student and tutor should strive to avoid Extended Miscommunication. When the tutor makes an important assumption about the student's intentions or beliefs, then the tutor should tell the student he is making that assumption. If the tutor had told student B that she assumed he was worried about the signs in the equations, they could have avoided several minutes of Extended Miscommunication. Alternatively, the students can take the lead in avoiding Extended Miscommunication by saying more explicitly what goals or beliefs they are using. Similar comments apply to avoiding the other non-productive activities: floundering, extraneous discussions, lectures and superficial kinds of "reflective follow up" or solution reviews.

Tutors too often tend to assume that because a student performs a correct action, they

have made correct inferences. For instance, in many problems, two objects are connected in such a way that they have the same acceleration. If the student produces for each body a Newton's law equation that has the same variable, a, for the acceleration, then tutors tend to assume that the student realized that the connection between the objects forces their accelerations to be the same. However, some students may not realize this—they tend to use a in all applications of Newton's law. At the risk of insulting the student, the tutor should say something like, "You are using the same variable for the accelerations of both blocks. Is that what you intended?" In other words, tutors tend not to do Coordination on occasions when the student is making Progress.

These are just a few, rather arbitrarily selected hypotheses that fall out of the model. It would be wise to think more deeply about the consequences of the model in order to find some counter-intuitive predictions that could be used to test the model.

Moreover, if the counter-intuitive hypotheses hold, then they might have the most practical importance, since they would influence tutors and students to change their practices in ways that they would not otherwise consider.

4 Plans for new studies

As described in the proposal for an extension to this line of work, we plan to conduct a main study of physics using human tutors, and two smaller studies, one using a simple computer-based physics tutor and the other using human cardiophysiology tutors. We plan to complete most of the study of human physics tutoring by the end of the current grant (April 30, 1996). In particular, we will have run the students, transcribed the protocols and completed most of the analyses. Finishing the analyses and writing up the results may take a bit longer. The work on the two smaller studies would start in mid-spring with data collection and protocol transcription; analysis would take place during the summer and early fall.

This section describes our current thinking about design of the main study. It should be considered only as a preliminary design. We would welcome comments from anyone who reads it.

Because we now believe that our pilot study was a success, we plan to use the same

basic design for the main study, but fix some of the problems and add a new type of analysis that couldn't be done in the pilot study because we had too few subjects.

Each student would engage in 6 activities:

1. Study physics (if they have not done so recently) up to the first chapter on Newtons law. This would include examples of problem solving but would not include any problem solving by the student.

2. Learn on how to use the OLAE user interface. (10 minutes)

3. Take a pre-test. (20 minutes)

4. Solve some problems with the aid of a tutor. (3 hours)

5. Take a post-test. (20 minutes)

6. Debriefing and commenting on how tutoring could be improved (10 minutes)

Activities 2 through 6 require 4 hours, so they will be split into 2 2-hour sessions. For efficiency, we would prefer to use subjects from University of Pittsburgh physics classes because we could skip activity 1. However, it is difficult to get enough of them into the lab *after* they have studied the appropriate chapter but *before* they do their homework for that chapter. Consequently, we may have to train (in activity 1) physics-naive but mathematically strong subjects from the Psych subject pool.

During the pre-test, tutoring and post-test, students will work on the OLAE user interface and give a verbal protocols. These will be our main data sources. We will insure that the pre-test, the post-test and the tutoring all address exactly the same content by conducting a rule-level analysis of the knowledge needed to solve each problem, which we can do with a combination of CASCADE, OLAE and hand analysis. The post-test and the pre-test should be equivalent, but the post-test has to be difficult enough that all subjects are not at ceiling. However, an overly difficulty pre-test can cause students to flounder, which would make it impossible to assess their competence on pieces knowledge that lay in parts of the problem space that they never reached. Thus, we will use tests that are partly problem solving and partly answering questions about portions of the problem solving process (e.g., "Please draw all the forces on block A." "Given the forces drawn in the figure,

write an expression for the x-component of each one."). Similar component-process tests were used successfully as part of the Chi et al. (1989) study.

We currently plan to conduct 3 analyses. Two are the successful ones used on the pilot data: (a) Locating learning events and finding their causal antecedents; (b) Parsing the protocols and perhaps even fitting them to a new version of the CASCADE model. However, because we will have more subjects, we will attempt to conduct a more traditional test of our hypotheses by seeking correlations between various types of tutoring episodes and learning outcomes. This analysis requires a bit more explanation.

As listed in Table 1 of the attached memo from Micki Chi, there are many hypotheses about what kinds of events during tutoring cause students to learn. For instance, negative feedback by the tutor and the subsequent discussion of the error might cause students to remedy defects in their knowledge. As a second instance, tutoring might increase the amount of self-directed knowledge construction, such as that found in during self-explanation (Chi et al, 1989) or Min analogy (VanLehn, 1995b). For each of these hypotheses (actually, we are revising and extending the list as we go), we are developing coding categories for the protocols. We will count instances of episodes of each category's. We will then run appropriate statistical tests to see which categories correlate strongly with the students' learning, as measured by their gain scores (pre-post test differences).

This third analysis is risky in that we may not have enough power to conclude anything from it. However, it we do get significant co-occurrences, then we will have demonstrated a relationship in a more direct way than the other two analyses.

5 Appendix: Illustrative parsing of a protocol

The problem statement is:

> Just as her parachute opens, a 670 N parachutist is falling at a speed of 50 m/s. After 0.80 s has passed, her speed has dropped to 11 m/s. Find the average retarding force exerted upon the parachutist during this time.

The correct solution is to note that there are two forces on the parachutist, her weight (acting downward) and the tug of the parachute (acting upward). The sum of these two

Table 1: Correct goal structure most closely matching the student's

```
solve problem
  qualitative analysis
    choose body (parachutist)
    identify forces acting on the body (weight & retarding force)
    identify the body's acceleration (upwards)
    draw a free-body diagram for the body
  quantitative analysis
    decide to apply f=ma to the parachutist vertically
      find mass
        using w=mg
          given weight
      find acceleration
        using a = (vf-vi)/t
          given vf, vi and t
          assume downward is positive
      find net force
        using sum of forces vertically
          given weight
          use variable for retarding force
          assuming downward is positive
      substitute into f=ma
    solve for retarding force
```

forces is equal to her mass (which can be calculated from her weight) times her acceleration (which can be calculated from the reduction in speed over the given duration). The ideal goal structure that most closely matches the one used by the student is shown in Table 1.

The following text provides a preliminary analysis of the protocol, illustrating the various types of tutoring shown in Figure 1. "S" stands for the student and "T" stands for the tutor. The numbers are line numbers.

Initialization 1-39: Starting up the software and instructions to student. This is not tutoring, so it is not included in the taxonomy.

Qualitative analysis 39-45: S constructs a qualitative analysis mentally, but doesn't draw a force diagram. She says virtually nothing during this time, despite the experimenter's prompts. The agency here is, of course, entirely the student's.

Start equation generation. This segment is coached problem solving. However, it ends with a debugging episode, which causes a shift in the focus of attention in the next segment.

46-48: S starts on quantitative means-end analysis (equation generation) by seeking an equation that contains the sought quantity, "retarding force."

47-56: S asks T for an equation that uses "the force that pushes up." She then clarifies the reference with "Is that still force N?" This illustrates the second type of agency, wherein a student reaches an impasse and seeks help.

57-91: T corrects terminological error. He tells S that the phrase "normal force" cannot refer to the retarding force because normal forces are caused by surfaces, that some forces do not have conventional names, so S can make her own name up, such as "retarding force." In passing, T asserts twice that the parachutist is decelerating. This episode is an instance of Debugging.

Qualitative analysis again This segment, from line 91 to line 174, is difficult to classify. On the one hand, S and T are doing some qualitative analysis together. Qualitative analysis is a bit disorganized, as there is not real internal goal structure to it. One just keeps making inferences until all applicable qualitative inferences have been made. On the other hand, there is a definite structure to this segment when viewed as a conversation. All the conversation refers back to the student's initial question. That conversational structure is not captured in the analysis. Also, it is not clear what the student's goals are throughout this section, and in particular, what led her to ask the initial question. Knowing her goal would help us classify this segment with more certainty.

92-100: S asks if "that force [the retarding force] equals the force going down?" It is not clear why S asks this question.

101-105: T supplies a co-referring phrase "You mean her weight?" S says "yes, her weight." S could learn that weight is a kind of force from this, but she later seems to use the rule intermittently later. This episode is an instance of Coordination.

106-115: T answers S's question. No, the forces are not equal because the parachutist is accelerating. S says, "Oh yeah, that's right." At this point, T has evidence that S did not successfully complete the qualitative analysis for the problem, because if she had, she

would not have asked the question. Therefore, T reopens the qualitative analysis phase with a combination of assertions (making the inferences himself) and questions (prompting S to make the inference).

116-120: T asserts the inference that the parachutist is accelerating because her speed is decreasing. Student says, "Right." This illustrates the type of agency where the tutor makes an inference without the participation of the student. If the tutor had involved the student, a learning event might have occurred.

121-138: T repeats a previous inference, that the acceleration means that the forces must not balance. Student says, "Right."

140-141: T asks about another inference in the qualitative analysis: Which direction is the acceleration? Tutors often ask leading questions like this one in order to set goals for the student. If such a question appeared in equation generation phase, it would have posed an equation generation goal that that an ideal student working alone would pose. However, in the qualitative analysis phase there is no strong goal structure. If one were working alone, one wouldn't first say "I wonder which direction the acceleration is going?" then proceed to achieve the goal of finding out. So we can't fairly classify the tutor's setting a goal in the qualitative analysis phase as the tutor being the agent of an inference that would normally occur anyway. This type of episode should probably be a new type of agency, but one that appears only in the qualitative phase.

142-162: S says "going down" and T debugs the incorrect answer.

163-174: T puts the upward acceleration result together with the imbalance result and infers "so there must be more force going up."

Drawing forces This segment is classic coached problem solving where S and T work together to finish up the qualitative analysis. (We do not include in the taxonomy the various user interface problems and the equipment breakage.)

175-215: T sets the goal for S of drawing the forces.

216-260: S has a problem with the user interface, which the experimenter fixes.

260-265: S draws a vector for what she calls "the parachutist's force...down." The user interface prompts her for a label. She is not sure whether to call the vector weight, 670N

or "the parachutist's force down." She tries to type help, but the user interface won't let her. She tries to ask the experimenter, but he won't answer her question (only the tutor answers physics questions).

266-369: The digital tape recorder breaks. T and the experimenter try to fix it, and give up, relying on the back up recorder.

370-372: S calls the vector "parachutist's downward force."

373-381: S draws a vector for the parachutist's speed. This might be prompted by the user interface, whose menu of vector types includes velocity.

382-413: S asks the experimenter if she can put numbers in the vector's labels. The experimenter says yes. She goes back and adds "50 m/s" to the velocity vector's label and "670N" to the force vector's label.

417-444: T interrupts and suggests only drawing forces because the immediate goal is to apply Newton's law. This is an episode of Debugging, prompted by a strategic error rather than a factual one.

445-452: S rereads problem statement to see if there are any more forces. She finds and draws "retarding force."

Equation generation again. This segment is coached problem solving that nicely illustrates the means-ends goal structure of Table 1.

453-454: S says, "now I am going to write an equation." The equation is probably $F = ma$, because that is the one the tutor mentioned twice earlier.

455-513: More trouble with the user interface.

514-535: S wants to apply $F = ma$, so she subgoals to find the acceleration. She knows the speed is dropping. She uses the appropriate kinematic equation ($a = (vf - vi)/t$), substituting numbers from the problem statement as she writes it, then uses a calculator to solve the arithmetic equation. This gives a negative number.

536-537: S decides to make downward be positive, because acceleration is upward and the number is negative.

538-539: S returns to $F = ma$, and subgoals to find the mass. She reaches an impasse, because she is confused about the distinction between mass and weight. This is the second

instance of this type of agency in the protocol.

540-567: S asks the tutor "should you convert the Newtons to kilograms?" T explains that 670N is a weight, and thus a type of force, so S should use $w = mg$. T does not explicitly correct her terminological error (using the concept of units conversion to refer to this calculation), but the Debugging is directed towards it as well as towards answering S's question.

568-625: S asks T to explain a contradiction between S's teacher and S's textbook. T tries to explain the difference between force and mass, but does a terrible job and stops in embarrassment. T concludes by tell S to apply $w = mg$ with $g = 9.8$. This is an instance of an Extraneous discussion.

626-644: S calculates mass from weight.

645-647: S returns to the goal of applying $F = ma$. She types and says "f=mg" but T does not correct the terminological error (using "g" when she means "a"). T probably didn't notice it.

648-659: S completes the application of $F = ma$ by substituting into it the quantities she derived for mass and acceleration.

Debugging. This segment illustrates two instances of Debugging, only one of which was necessary.

660-708: S starts to enter the results ($-3330N$) as the answer. T interrupts with negative feedback. T says that the "F" in $F = ma$ stands for the sum of forces. He concludes, "so you need to have both of those forces equal to your 3330."

709-733: S asks about the T's use of the term "3330" by asking, "isn't it negative, though? Since downward is positive?" From S's point of view, this was an instance of Coordination, because she thought that T was using an abbreviated reference to her formula $-3330N$. T should have just answered "yes." However, T has not noticed that S has used downwards as positive in her use of the negative sign with both acceleration, the net force and now her question. He states that because upwards is positive, the net force should be positive. Thus, from T's point of view, this episode is Debugging.

734-759: More interface problems.

Equation generation again. During this segment, S and T return to goal of generating equations. Although S had thought she was finished and thus presumably emptied her goal stack (memory), she has no trouble resuming problem solving.

760-775: S sets the net force equal to the two forces, gets their signs right with the assumption of upwards being positive, and solves for the retarding force.

776-789: T congratulates S on a correct solution. S thanks T for reminding her about F being a sum of forces.

Recap. 790-821: T reviews the whole problems solution. S just says "right" whenever he pauses. This student's session occurred early in the study. T was trying to do a "reflective follow-up" like the ones used successfully by the Sherlock and Smithtown tutoring systems. Such recaps never seemed to cause learning, nor even hold the student's interest, so T stopped doing them.

References

Baker, M. (1994). A model for negotiation in teaching-learning dialogs. *Journal of Artificial Intelligence and Education*, 5(2):199–254.

Chi, M., Bassok, M., Lewis, M., Reimann, P., and Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 15:145–182.

Chi, M., Glaser, R., and Rees, E. (1982). Expertise in problem solving. In Sternberg, R., editor, *Advances in the Psychology of Human Intelligence, Vol. 1*, pages 7–75. Lawrence Erlbaum Associates, Hillsdale, NJ.

Chi, M. and VanLehn, K. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences*, 1:69–105.

Newell, A. and Simon, H. A., editors (1972). *Human Problem Solving.* Prentrice-Hall, Inc., Englewood Cliffs, NJ.

Siegler, R. S. and Jenkins, E. (1989). *How Children Discover New Strategies.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Ur, S. and VanLehn, K. (1995). Steps: A simulated, tutorable physics student. in press, Journal of AI and Education.

VanLehn, K. (1991). Rule acquisition events in the discovery of problem solving strategies. *Cognitive Science*, 15(1):1–47.

VanLehn, K. (1995a). Cognitive skill acquisition. To appear in Spence, J., Darly, J. and Foss, D.J., editors, *Annual Review of Psychology, Vol. 47.*

VanLehn, K. (1995b). Looking in the book: The effects of example-exercise analogy on learning. Submitted for publication.

VanLehn, K. (1995c). Rule learning events in the acquisition of a complex skill. Submitted for publication.

VanLehn, K. and Jones, R. (1993a). Integration of analogical search control and explanation-based learning of correctness. In Minton, S., editor, *Machine Learning Methods for Planning*, pages 273–315. Morgan Kaufmann, Los Altos, CA.

VanLehn, K. and Jones, R. (1993b). Learning by explaining examples to oneself: A computational model. In Chipman, S. and Meyrowitz, A., editors, *Cognitive Models of Complex Learning*, pages 25–82. Kluwer Academic Publishers, Boston, MA.

VanLehn, K., Jones, R. M., and Chi, M. T. H. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, 2(1):1–59.

Summary of Potential Analyses of Tutoring Protocols

Micki Chi

Nov. 3, 1995

The overarching question addressing tutoring research is why tutoring is so effective. There are three broad foci. One can attribute the enhancement of learning to the role of the tutor. That is, what is it that a tutor does during tutoring that makes learning so effective. Alternatively, one can attribute the enhancement to what the student gets to do in tutoring situations. Finally, one can attribute the learning gain to the interaction of the tutor and the tutee.

Within each of these broad factors, we can scrutinize on several subfactors. With respect to the first, the tutor's role, one can attribute the learning gains to several possible subfactors. First, it could be that tutors are accurate and able to assess tutee's missing knowledge and diagnose their misunderstanding (more about this later). Second, tutors can be very didactic, and give away much information that tutees need through extensive explanations . Third, tutors have lots of opportunities to give feedback, both corrective and reinforcing ones. (Corrective feedback means telling the tutee what s/he has done is wrong, and reinforcing feedback means telling the tutee that what s/he has done is right.). Fourth, tutors can scaffold tutees. For example, instead of giving direct explanations, tutors can instead direct the tutee, decompose a problem into subcomponents for the tutee, point out salient features, etc. Fifth, tutors can be effective because they prompt students to construct meanng and self-explain (see below). Finally, tutors can be motivating.

With respect to the second factor, the student's role, one can attribute the learning gains to several subfactors as well. First, tutees in a tutoring situation have more opportunities to self-explain and construct meaning. Second, tutees have more opportunities to ask questions. Although asking questions is a form of construction, asking question is more specific than merely constructing. It is pursuing explicitly aspects of the content that the tutee fails to understand. As Graesser's work notes, tutoring affords the much needed opportunites to ask questions. Finally, the tutee can more explicitly follow and carry out the tutor's plan, whereas in the classroom context, the tutee may not be actively carrying out the tutor's plan.

With respect to the third factor, it may be that a certain type of interaction between the tutor and the tutee that leads most effectively to learning. For example, a tutee might learn the most when a tutor has correctly diagnose his/her misunderstanding and give explanations about it. Or alternatively, a tutee might learn the most when a tutor scaffolds her only and gives her the opportunity to construct meaning, without explicit diagnosis of her misunderstanding. Or alternatively, a tutee learns only when corrective feedback is given when it is explicitly requested. Or possibly, by giving corrective and reinforcing feedback, a tutor has a greater chance of success at making the tutee follow his plan.

Finally, it is possible that interactions can occur within the tutor's own actions. For instance, an accurate diagnosis of student's misunderstanding will allow the tutor to generate better didactic explanations or to give more accurate reinforcing feedback (accurate means giving reinforcing feedback at the appropriate time).

All these factors are summarized in Table 1. Thus, Table 1 represent what I think a priori is going on in a tutoring context. I now explain how the protocols are coded; give some preliminary result of one problem from one subject, and give pointers about how a more careful, thorough, and extensive analyses can be further carried out.

How Protocols Are Coded

The protocols are basically coded according to segments that correspond to an *action* undertaken either by the tutor or the tutee. The ineresting/important actions correspond more or less to the ones listed in italics in Table 1. Table 1 lacks many actions that are actually used, such as *reviews* , *summarizes* , *reminds, analogizes*, and *responds* as used by the tutor, and *calculates, applies, confuses,* as used by the tutee. (The actions in Table 1 may have to be revised again to incorporate all or some of the ones that were actually used in the coding.) Thus, responses such as "Oh" or "Okay" are not coded. Thus, the result of the first pass of the protocols is an annotated version with the actions coded. Several iterations had to be done to get this first pass result. That is, each iteration refined the actions that need to be coded.

The second pass at the protocols consisted of identifying the pieces of knowledge that the tutee either lacked or was confused about. At present, for Problem 1, I have identified 9 pieces of knowledge that can be learned. There are several differences between these knowledge pieces, even though for the time being I have simplify them by listing them as 1 to 7a (See Table 2). Let me note several things about these knowledge pieces.

a) First, the majority of these knowledge pieces are general (i.e., cuts across different physics problems, such as confusing mass with weight, Item 1 in Table 2), but there are a small subset that is generated from the specific problem at hand. The example is not as clearcut in this subject's data (Subj. A), but it is more apparent in others'. One could say, perhaps, that Knowledge Piece 6 (KP6), confusion about the signs, is specific to this problem, although this can be a general confusion. A better example comes from Subj. Q, who confuses what she calls retarding force for Problem #1 with normal force. So, for Subj. Q, only 1 out of 9 KP are specific to this problem. For Subj. A, none of her 9 Kps are specific to this problem.

b) For two subjects anyway, there seems to be a co*mmon set of knowledge pieces that are* misunderstood by them. So, for Subj. A and Subj. Q for Problem #1, they both are confused about KPs 1 through 6, including 3a (or 9).

c) The majority of the knowledge pieces that I have identified initially from the protocols arise from the tutee's confusion. These are KP 1-6 (including 9), but not 7 or 7a. Henceforth, I will call these misunderstandings or misconceptions.

d) Occasionally, the tutor realizes that the tutee needs a knowledge piece that the tutee has not been aware that it is missing, and the tutor simply tells the tutee these knowledge pieces. A good example is KP7a, where the tutor tells the tutee the rule that IF something is speeding up, THEN there is acceleration. I will call these *missing* knowledge pieces as opposed to *misconceived* knowledge pieces, as explained in Item c above. For Subj. A in Problem #1, there is only one missing KP of this kind. Missing KP has the property that they are pretty straightforward and can easily be put into condition-action rule terms.

e) Fifth, although a careful count has not been made, all the misconceived KP occur at more than one occasion in the protocols, whereas the telling of missing KP occurs more or less once. In other words, I am very confident that these are the misconceived KPs.

f) Sixth, there is one other missing KP, and it is unique in that it is not only procedural in nature, but it encompasses all the other KPs plus more. I am talking about KP7, which I included later. It is a procedure of "drawing forces on the diagram" that is initiated by the tutor, for which the tutee lacks. But drawing the forces incorporate knowledge about finding them, and realizing such things as that weight is a force but not mass (KP1), or that gravity contributes toward the weight (KP5), etc. To put it in another way, Kps 1-6 are confusions exhibited by the tutee, whereas KP 7 and 7a are missing KPs that the tutor wants the tutee to learn.

g) Perhaps the most important thing to note about these Kps (1-6) is that they are subtle misunderstandings that are not explicated by the text or by the tutor in any explicit or direct way. (In fact, it took me a long while to figure them out.) The two that are most salient in this respect are Kps 3a and 3b (or 4 and 9). They are subtle but important variation of KP3. That is, KP3 simply states the syntactic form that the F in F=ma in Newton's Law specifies the *sum* of forces. However, there are two corollaries to this KP3. KP3a or KP9 is the realization that an unknown, to be sought, can be one of these forces, and it may not have a name. This is related to KP4 (or KP3b). That is, students seem to think that an unknown force is one that has a name, and can be set as a variable, and there ought to be an equation that corresponds to that variable. Thus, they have trouble thinking of an unknown force simply as another force that has to be added to the same equation. This kind of confusion is exhibited by remarks or confusions about not knowing how to find the friction force (since they don't know the friction equation), instead of treating friction force as just another force that can be summed in the F's.

Tutee's Learning

What Did the Tutee Learn

During the third pass through the protocols, I look for evidence *to see which knowledge* pieces are learned. There are three categories: yes, no, and maybe (shown by a question mark in Table 2). (I have not systematically worked through this third pass again in order to get an operational definition of how I deduce that a KP is learned.) Basically, a KP is coded as learned if the tutee exhibited the correct application or use of that KP in relation to how it was used or conceived of before (it seems subjective, see next comment below, but I can probably come up with more operational reason if I go over it again).

For example, for KP1, the tutee was initially confused about weight and mass in L133 where she says explicitly that "I guess since it's falling downward then it's the mass would equal the weight.", and then again in L 153 when she explicitly uses F=ma by substituting the weight of 670 N for the mass. Thus, there are at least two occasions when this confusion was initially exhibited.

It is apparent that she removed this confusion not only because she applied it in the equation correctly in L383 "For the mass...um, you wanted me to take 670 and divide it by

9.8?" and again in L391 "Um, well, yeah, that's the weight and I need to find mass". But she also realized that she had errored before, in her comment in L371 "I know he was saying about the mass...I actually took the weight,...". Thus, I am certain that she learned KP1 because there are three instances of getting the correct conception and applying it correctly in an equation form.

So, out of a possible of 9 KPs, she learned 3 at the end of Problem #1. A KP is categorized as not learned (the "no" category), if either the following is true. First, the tutee could have exhibited a lack of undertanding in some later portion of the protocols of this problem even though she could use the Kps correctly earlier in the protocols. These early occasions usually arise from following the tutor's instruction closely, so that one is hard pressed to say that she understood and learned it. Hence, the evidence that she failed to have learned it arises from manifestations of misunderstanding again at a later portion of the same problem. Second, we know the tutee has not learned it because the KP is confused again in a later Problem #2. Four of the 9 KPs are not learned.

The "maybe" categories are those KP for which we have no definitive evidence that the tutee has either learned it or not. In all these cases, the tutee followed the tutor's instruction in the application and use of the KP. And often because there are no further occasions to use or exhibit knowledge of this KP, we cannot ascertain whether in fact it was learned or not, hence, the maybe category. There are two "maybe".

How Did the Tutee Learn It

This analysis looks at those KPs that were learned and see which type of tutorial interactions resulted in the learning. Since there are only 3 Kps learned, we can go over each of them. KP1 was learned by the tutor's prompt/hint. The tutor merely warned "But the mass might give you a little trouble."

KP5 (that weight is a force due to gravity) was learned by the interchange occuring in L387-L402. This is learning occuring from the interaction of the tutee with the tutor. In this case, the tutee asked the tutor directly about her confusion in L392 "but why don't I need to take the, um, the gravity into account, though?" and the tutee learned from the response given by the tutor in L400 and again in L405.

KP7a, that speeding up means there is acceleration, is misunderstood by the tutee in the context about the direction/sign of the acceleration, which is related to this KP. So the tutor corrects the direction of the sign by telling the tutee in L 451 and L456 that "IF it's accelerating downward (taking it as positive), THEN that would mean she's speeding up." The reason I am sure that she learned this KP is because she was able to use it and rephrase it as "IF it is *slowing down*, THEN the acceleraton is decreasing."

In sum, KPs are learned either by the tutor's prompt, answers to tutee's explicit questions, or explicit feedback on something that the tutee did that was wrong. No learning resulted from tutor's actions such as diagnosing misconceived knowledge, giving extensive didactic explanations, etc.

Tutor's Actions

Does the Tutor Diagnose?

This section of the analyses focuses on which tutor actions seem most facilitative to learning, and more generally, which tutor actions seem to dominate. The one that I will focus on is whether the tutor diagnoses the student's misunderstanding or not, since this seems to be an unresolved issue in the literature. First, let me clarify some issues. I distinguish between diagnosing student's misunderstanding from assessing missing knowledge. Diagnosing student's misunderstanding is when a student clearly exhibited a misunderstanding, but the tutor ignores it rather than try to either figure out more deeply what that misunderstanding is, or redressing it. Out of 41 tutorial actions coded for Problem #1, there were at least 6 occasions for which the tutee manifested confusion. In 5 of these cases, the tutor ignored it. For example, in L133-154, the tutee was confused about mass and weight. This confusion was never resolved until much much later. Instead, the tutor tends to pursue his own plan of making the tutee find all the forces. This is an instance of failing to diagnose student's misunderstanding because an obvious misunderstanding was exhibited, and the tutor ignores it rather than clarifies it. Another really good example comes from Problem #3. There, the student was exhibiting confusion about which direction is the normal force; that is, is it the force from the bottom block on the top block or vice versa. The tutor simply ignored this confusion for a long time.

In contrast, assessing missing knowledge is when the tutor, from prior experiences in tutoring or teaching, realizes that either certain concept is difficult or some aspect of the problem may present difficulties. In such a case, the tutor assesses whether the tutee knows this piece of knowledge even though the tutee may not have exhibited any lack of knowledge or understanding. A good example of this assessment of missing knowledge occurs in l247-251 where the tutor plunges ahead and tell the student to beware of the sign even though there is no evidence that the tutee would be confused about the signs. This is because Subj. A is the second subject, and the tutor experienced Subj. Q's difficulties with the signs.

A second discrimination I want to make is between a failure to diagnose versus an unawareness of the tutee's confusion. There are at least 2 occasions in Problem #1 where the tutee was confused but the tutor was completely oblivious to it.

What Does the Tutor Do?

If the tutor is not diagnosing and addressing tutee's misunderstanding, what is it that the tutor does? The tutor mostly pursues his own plan. There are two related subplans within his plan. First, the tutor wants the tutee to draw all the forces on the diagram, and then sum the forces. The tutor executes his plan at a fairly high level, and resorting to a more detailed level only when the tutee requests information and fail to execute any actions.

The tutor also gives feedback. Here I distinguish between two kinds. One kind is *corrective feedback*. That is, feedback that corrects something that the tutee did that was wrong. Corrective feedback tends to occur mostly in syntactic actions (this needs to be quantified and verified). For instance, when the student adds some wrong forces together, or when mass is substituted for weight in the equation, the tutor will interject and give corrective feedback. In contrast, the tutor seldom gives corrective feedback when there is a conceptual misunderstanding (which I have coded above as a failure to diagnose).

The tutor gives *reinforcing feedback* when some correct action was taken, usually in equation form or in drawing of forces on the diagram. Here, even though the tutee is not yet sure that she is doing it correctly, the tutor interjects and says it's correct. (This is another form of ignoring tutee's understanding; but this is not counted as such in the above analyses on ignoring tutee's misunderstanding.)

There really wasn't that many occasions of prompting and scaffolding, partly because of the nature of this tutoring task (problem solving), which tends to be very scripted in getting the right equation down.

Some Conclusions

If I had to say how learning takes place in tutoring context or why it is so effective, I would attribute it to the following factors. First, the student ultimately learns how to solve the problem because the tutor insists on her following his prescripted plan. That's not to say that she has learned that much. But by following a prescripted plan a couple of times in a tutorial session, a tutee will pick up the right procedures and such.

The knowledge pieces, on the other hand, are learned not by following a prescripted plan, but rather, by the opportunities for the tutees to ask explicit and specific questions. If confusions are exhibited generally, the tutor tends to ignore it. But if the tutee explicitly asks a question, then the tutor will usually answer it. From these answers, the tutee can learn. The tutee, on the other hand, does not usually learn from long-winded didactic explanations, possibly because these explanations do not address her misunderstanding appropriately. (This conclusion can be quantified by counting how many times a tutor has to repeat some didactic explanation.) Thus, the tutee is learning from interaction with the tutor.

The tutor is definitely not diagnosing the student's misunderstanding; some times he is even unaware of them. The fact that he reinforces correct actions without ascertaining the student's certainty about that action is more support for this conclusion. (This can be documented and quantified as well.) Also, because the tutor is not diagnosing the tutee's misunderstanding, there is no learning as a result of the interactions among the tutor's own actions (see Item D in Table 1).

There may be drastic differences between tutoring practice that is focused on solving problems correctly, versus tutoring practice that is aimed at understanding some concepts. In the latter case, the tutor may engage in more scaffolding and prompting.

Additional Anslyses

What I have done is give you an idea of what can be gleaned from these protocols. The conclusions drawn here are based on one subject on one problem. For Subj. A, I have identified a total of 13 KPs across the 3 problems; and all the protocols for the 3 problems are coded. Two of Subj. Q's protocols are also coded. Certainly the entire protocols need to be processed several more times to get them all analyzed. Moreover, many other conclusions I suggested can be quantified as well.

Table 1

Factors Responsible for the Enhancement of Learning During Tutoring

A. Tutor's Role

l) Accurate at *assessing* missing knowledge

2) Accurate at *diagnosing* misconceived knowledge

3) Opportunity to give extensive didactic *explanations*

4) Opportunity to give *corrective feedback*

5) Opportunity to *reinforce* correct moves

6) Opportunity to *scaffold* tutees

7) Opportunity to *prompt* tutees

8) Opportunity to *motivate* tutees

9) *Query* tutee in the context of tutor's plan

10) *Tells* tutee the next step in tutor's plan,...

B. Tutee's Role

1) Opportunity to *self-explain* and construct meaning (including generating inferences, making connections between what is currently known with previous knowledge, following own plan, etc.)

2) Opportunity to *ask questions*

3) *Following* Tutor's plan

C. Potential Enhancing Interaction between Tutor and the Tutee

1) Tutor's prompting leads to more self-explanations

2) Tutor's scaffolding leads to more self-explanations

3) Tutor giving corrective and reinforcing feedback allows Tutee to follow Tutor's plan

D. Potential Enhancing Interaction between Tutor's Own Actions

1) Accurate diagnosis of misunderstanding lead to better didactic explanation
2) Accurate diagnosis lead to better reinforcing feedback; .....

#1

(yes/no/?) Is assessment that S has learned it or not after the lst problem (lst column), 2nd problem (2nd column). ? Means they complied with T's instruction, but not sure they got it on their own.

Table 2

yes	1) confuses m with wt (L132, 154)
?	2) confuses direction of motion with direction of acceleration(L410-465)
no	3) does not realize that F=ma is sum of forces==>
no	9) or 3a) that an unknown force f can be one of the forces embedded in F for the sum of forces
no	4/3b) does not realize that all forces acting on a body can simply be summed, without necessarily needing to know exact equation involving that force (T tells L225)
yes?	5 or 1a) Does not realize wt is force due to gravity (L387)
?	6) confusion about signs (T alert L258; L420)
no?	7) draw forces on diagram (A procedure)
yes	7a) If slowing down/speeding up, means there is acceleration.(L448; learned in 465)

Table 3

 # 2

yes? 8) The system one focuses on is the system that receives the forces, not the agent of the force.

? 9) Can't conceive of forces that are not named/or an unkonw force, like force of smaller2 kg block pushing on larger block, as a force acting on a system, but can only treat it as a syntactic unknown variable, there never draws an unknown force on the diagram

10) IF 2 blocks are connected and moving, the acceleration of one of the blocks is the same as the accel of the 2 blocks together.

11) IF there are not other point of contact, THEN there are no more contact forces. (See L878)

12) IF a force is at a rt angle to direction of motion, THEN can ignore it (L888)

 # 3

13) Doesn't know whether the normal force is one pushing from bottom to top or top to borrom

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
11/16/95

3. REPORT TYPE AND DATES COVERED
Progress Report; 5/1/94 - 11/16/95

4. TITLE AND SUBTITLE
Progress Report: Towards a theory of learning during tutoring

5. FUNDING NUMBERS
N00014-94-1-0674

6. AUTHOR(S)
Kurt VanLehn
Michelene T.H. Chi
William Baggett & R. Charles Murray

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Learning Research and Development Center
3939 O'Hara St.
University of Pittsburgh
Pittsburgh, PA 15260

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Navy, Office of the Chief of Naval Research
800 N. Quincy St., code 1511:LG
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This progress report has 4 sections. The first gives an overview of the project, our pilot study, and our plans for future work. The second and third sections present the results from the pilot study in more deatil. The last section discusses our plans for the main study.

14. SUBJECT TERMS
tutoring, OLAE

15. NUMBER OF PAGES
38

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	unlimited